PROJECT REPORT

Course:EE673

Submitted By:-

Kritarth Mishra-2020CEB1016

Kritika Garg-2020EEB1180

VIOLENCE DETECTION

Title of the Base Paper: CNN-BiLSTM Model for Violence Detection in Smart Surveillance

Venue: © Springer Nature Singapore Pte Ltd 2020

ABSTRACT

- **Objective:**

1. To create a simple computer program to tell apart violent and non-violent actions better.
2. To help deal with more fights happening in public places because of different social and money issues.
3. Make it easier for government and public workers to catch these fights using smart cameras.
4. Make a smart system that can quickly spot violent behavior in real-time.

- **Methodology:**

Temporal and Spatial Features:

1. To classify violent or non-violent actions, our model needs to predict patterns in consecutive frames, considering both the movement of subjects and their degree of motion.
2. Temporal features, related to time, must be considered alongside spatial features to detect sequences in frames accurately.

CNN:

1. The model includes a Convolutional Neural Network (CNN) with layers for convolution and max pooling.
2. Each convolutional layer uses 64 kernels with a 3x3 kernel size and passes through a "relu" activation function.
3. Max pooling is performed with a filter size of 2x2.
4. TensorFlow and Keras API are used for deploying the CNN.

Bidirectional LSTM Cells:

1. Long Short Term Memory (LSTM) cells are employed to reconsider previously trained features.
2. LSTM cells help in remembering past events and are capable of working in both forward and reverse directions.
3. A bidirectional LSTM layer is created by combining mechanisms from both directions, enhancing data storage accuracy.
4. Bidirectional LSTM compares frame sequences in both forward and reverse directions, adding robustness to the model for violence detection.

Dense Layers:

1. Fully connected dense layers, a common feature in deep learning, are utilized to extract features.
2. The features are flattened and forwarded to the next model.

Dataset:

The CNN-Bidirectional LSTM model is evaluated on standard datasets for violent and non-violent action detection, including Hockey Fights, Movies, and Violent Flows datasets.

Data Preprocessing:

1. Frames are extracted from videos and reshaped to 100x100 pixels.
2. Training data are organized into sequences, with each sequence representing a pattern in the video.
3. Numpy arrays are used to handle the training data, with each row representing a sequence of frames.

Training Methodology:

1. A group of 10 consecutive frames is fed into the model to extract spatial and temporal features.
2. Stochastic gradient descent is used as an optimizer with a learning rate of 0.01 and decay of 1e-6.
3. "Sparse categorical cross entropy" is employed as the loss function for the multi-class classification problem.
4. The dataset is divided into a 9:1 ratio for training and testing, and the model is trained for 25 epochs to maintain computational efficiency.

- **<u>Findings:</u>**

1. The program can quickly and accurately spot fights as they happen.
2. The program works better than other ways of finding fights.
3. Putting this program into smart cameras can help watch for fights automatically.
4. This program can help keep people safer by stopping fights before they get worse.

<u>Contribution of the paper:</u>

1. New Model Introduction:

   Introduces a fresh CNN-BiLSTM model for spotting violent and non-violent actions in videos.

2. Better Performance:

   Shows improved results compared to existing methods on standard datasets, proving the effectiveness of the new model.

3. Consideration of Time Context:

   Considers both past and future movements in videos, making it better at predicting and pinpointing violent events.

4. Potential for Prevention:

   Suggests the model could evolve into a tool for preventing violence by analyzing past events, offering possibilities for proactive safety measures.

<u>Proposed Novelty:</u>

1. **Batch Normalization:(I have added it after every layer)**

- Stabilizes the training process.
- Speeds up training by normalizing activations.
- Improves generalization ability by reducing internal covariate shift.
- Adds stability to gradients throughout the network.

2. **Learning Rate Scheduling:(Added in a separate code)**

- Dynamically adjusts learning rate during training.
- Utilizes techniques like ReduceLROnPlateau or learning rate decay.
- Helps in faster convergence.
- Can lead to better performance by fine-tuning parameters effectively.

3. **Attention Mechanism:(added along with batch normalization as a self)**

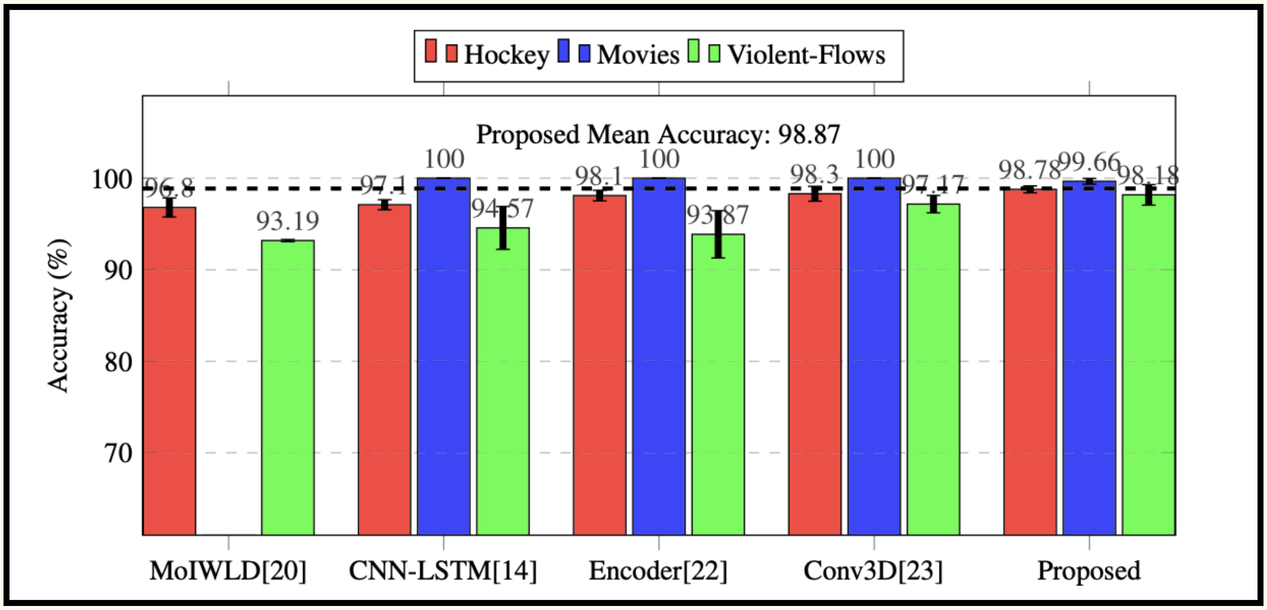
- Integrates attention mechanism into the model.
- Allows focusing on relevant parts of input data.
- Particularly useful for identifying violence by weighing frames or features differently.
- Enhances the model's ability to capture contextually relevant information.

<u>Results:</u>

(We ran the code for hockey dataset only as you can see from provided code)

The model achieves impressive classification accuracies: 99.27% for Hockey Fights, 100% for Movies, and 98.64% for Violent-Flows datasets.

In comparison to other models:



Instructions for running the code:

1.)Here is the drive link for data in which hockey data set it contained for "hockey_fights_normal.ipynb"

https://drive.google.com/drive/folders/1uVBz03RoCJzfuhfCM8FWIvE4gAFuvVvp?usp=share_link

2.)For "hockey_fights_automatic" data is automatically extracted.

3.)Provide the respective paths of data at appropriate places.

4.)Make a dataframe named folder.

5.)Provide its path for data extraction.

6.)Run the code.

7.)You can see model summary and accuracy measures from the report itself.